

A Regression Framework for Rank Tests Based on the Probabilistic Index Model

Jan De Neve and Olivier Thas

We demonstrate how many classical rank tests, such as the Wilcoxon–Mann–Whitney, Kruskal–Wallis and Friedman test, can be embedded in a statistical modelling framework and how the method can be used to construct new rank tests. In addition to hypothesis testing, the method allows for estimating effect sizes with an informative interpretation, resulting in a better understanding of the data. Supplementary materials for this article are available online.

Keywords: Friedman rank test; Kruskal–Wallis rank test; Probabilistic index; Wilcoxon–Mann–Whitney test

*Jan De Neve is postdoctoral researcher at the department of Mathematical Modelling, Statistics and Bioinformatics of Ghent University, Coupure Links, 653, 9000 Gent, Belgium (E-mail: JanR.DeNeve@UGent.be). Olivier Thas is associate professor at the department of Mathematical Modelling, Statistics and Bioinformatics of Ghent University, Coupure Links, 653, 9000 Gent, Belgium and honorary professor at the National Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia (E-mail: Olivier.Thas@UGent.be). The authors gratefully acknowledge the Research Fund Flanders (FWO) research grant G020214N and the IAP research network grant no. P7/06 of the Belgian government (Belgian Science Policy).

1 INTRODUCTION

The Wilcoxon–Mann–Whitney (WMW) ([Mann and Whitney, 1947](#); [Wilcoxon, 1945](#)) and Kruskal–Wallis (KW) ([Kruskal and Wallis, 1952](#)) tests are well known and popular rank tests to analyze two- and K -sample layouts. These rank tests are distribution-free, robust, intuitively appealing, and do not necessarily focus on the mean response. For the WMW test, for example, the alternative hypothesis is expressed in terms of the probability $P(Y_1 < Y_2)$, where Y_1 (Y_2) denotes the continuous response of the first (second) group. It is the probability that a random observation of the second group exceeds a random observation of the first group. The alternative hypothesis of the KW test can be expressed using similar probabilities, $P(Y < Y_k)$, where Y_k denotes the response in group $k = 1, \dots, K$, and Y the response associated with the marginal response distribution; see e.g. section 9.6.1 in [Thas \(2009\)](#). It is the probability that a random observation of group k exceeds a random observation of the marginal distribution. Under additional assumptions, such as location-shift, the alternative can also be expressed in terms of means or medians, and for a given family of distributions, locally most powerful rank tests may be constructed for testing equality of means; see e.g. [Hajek et al. \(1999\)](#); [Lehmann \(1998\)](#).

Following the introduction of the first rank tests for the two- and K -sample layout, a vast number of rank tests for more complicated designs have been developed. However, rank tests are typically constructed for specific designs making it difficult to extend them to customized designs. Furthermore, rank tests for more complicated designs are often not implemented in software packages, thereby hampering their widespread use.

For classical parametric tests with focus on the mean response, such as the two-sample t - or ANOVA F -test, this barrier is circumvented because they arise naturally from the General Linear Model (GLM) framework. Hence, for more complicated

designs the correct hypothesis test results from formulating the appropriate GLM. Basic knowledge on GLMs is often sufficient for analyzing data from a variety of designs and the GLM is available in most statistical software packages.

In this paper we situate a large class of rank tests within the semiparametric Probabilistic Index Model (PIM) framework of [Thas et al. \(2012\)](#). This model can be seen as the rank-equivalent of the GLM, but not to be confused with the rank-transform approach of [Conover and Iman \(1981\)](#). We will show that a transformation to binary *pseudo-observations* is more flexible than the rank-transform. Furthermore, by embedding the method in the PIM framework we can relate the tests to parameters with a well defined interpretation on population level. The full connection between a PIM and rank tests were not exploited in [Thas et al. \(2012\)](#) since specific parametrizations are required. More specifically, depending on this parametrization, we establish a simple connection between the PIM and the WMW, KW, Friedman ([Friedman, 1937](#)), Mack–Skillings (MS) ([Mack and Skillings, 1980](#)), Brown–Hettmansperger (BH) ([Brown and Hettmansperger, 2002](#)) and the Jonckheere–Terpstra (JT) ([Jonckheere, 1954](#); [Terpstra, 1952](#)) rank tests and some tests proposed by [Brunner and Puri \(2001\)](#) based on the work of [Akritas and Arnold \(1994\)](#).

The PIM framework also allows developing new rank tests and in addition to hypothesis testing, the PIM parameters possess an informative interpretation as effect sizes.

Instead of modelling the conditional mean, a PIM models the probability $P(Y_i \preceq Y_j) := P(Y_i < Y_j) + 0.5P(Y_i = Y_j)$, where Y_i and Y_j are independent response variables with distribution functions F_i and F_j , respectively. This probability is known as the *probabilistic index* (PI); see, for example [Acion et al. \(2006\)](#); [Thas et al. \(2012\)](#) and the references therein. The distributions F_i and F_j may depend on p -dimensional regressor patterns, say \mathbf{X}_i and \mathbf{X}_j . Let (Y_i, \mathbf{X}_i) , $i = 1, \dots, n$, denote the i.i.d. sample

observations, then a PIM is defined as

$$P(Y_i \preceq Y_j \mid \mathbf{X}_i, \mathbf{X}_j) = m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}) = g^{-1}(\mathbf{Z}_{ij}^T \boldsymbol{\beta}), \quad (\mathbf{X}_i, \mathbf{X}_j) \in \mathcal{X}_n, \quad (1)$$

where \mathcal{X}_n is the set of pairs of regressors $(\mathbf{X}_i, \mathbf{X}_j)$ for which the model is defined. For most models in [Thas et al. \(2012\)](#), the PIM is defined for all n^2 pairs so that $\mathcal{X}_n = \{(\mathbf{X}_i, \mathbf{X}_j) \mid i, j = 1, \dots, n\}$. We will show that the connection with rank tests sometimes requires that the scope of the PIM is restricted to a subset of pairs. For the randomized complete block design of [Section 3.2](#), for example, the PIM is only defined for pairs of observations within the same block so that the responses in $P(Y_i \preceq Y_j \mid \mathbf{X}_i, \mathbf{X}_j)$ refer to outcomes in the same block.

The function $m(\cdot, \cdot; \boldsymbol{\beta})$ has range $[0, 1]$ and if $(\mathbf{X}_i, \mathbf{X}_i)$, $(\mathbf{X}_i, \mathbf{X}_j)$, and $(\mathbf{X}_j, \mathbf{X}_i)$ are elements of \mathcal{X}_n , then $m(\cdot, \cdot; \boldsymbol{\beta})$ must also satisfy $m(\mathbf{X}_i, \mathbf{X}_i; \boldsymbol{\beta}) = 0.5$ and $m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}) + m(\mathbf{X}_j, \mathbf{X}_i; \boldsymbol{\beta}) = 1$. Similar to generalized linear models, the function $m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta})$ is restricted to $g^{-1}(\mathbf{Z}_{ij}^T \boldsymbol{\beta})$, where $g(\cdot)$ is a link function and \mathbf{Z}_{ij} depends on the regressors \mathbf{X}_i and \mathbf{X}_j . For a general \mathbf{Z}_{ij} and $g(\cdot)$, [Thas et al. \(2012\)](#) provide a semiparametric consistent and asymptotically normally distributed estimator of $\boldsymbol{\beta}$ and a consistent sandwich estimator of its asymptotic variance. They primarily consider $\mathbf{Z}_{ij} = \mathbf{X}_j - \mathbf{X}_i$ together with the logit and probit link function. In this paper we show that different choices of \mathbf{Z}_{ij} and $g(\cdot)$ are required to establish the relationship with rank tests.

The paper is organized as follows. In [Section 2](#) we give the estimating equations related to PIM [\(1\)](#) and summarize the most important distributional properties of the estimator. In [Section 3](#) we propose a parametrization for factorial designs and establish the connection with the KW, Friedman, and MS tests and illustrate how they can be extended to the two-way layout. In [Section 4](#) we consider a second parametrization and demonstrate the connection with the WMW, BH, and JT tests. In [Section 5](#) we show how a more powerful rank test than the WMW test can be constructed upon incorporating baseline information of a covariate. In [Section 6](#) an

example dataset is analyzed to illustrate the potential of a regression framework for rank tests and Section 7 gives the conclusions and a discussion in which we relate our approach to other methods.

2 ESTIMATING EQUATIONS AND DISTRIBUTIONAL THEORY

Let \mathcal{I}_n denote the set of indices (i, j) for which $(\mathbf{X}_i, \mathbf{X}_j) \in \mathcal{X}_n$ and define the *pseudo-observations* $I(Y_i \preceq Y_j) := I(Y_i < Y_j) + 0.5I(Y_i = Y_j)$ with $I(\cdot)$ the indicator function. An estimator of $\boldsymbol{\beta}$ in (1), say $\hat{\boldsymbol{\beta}}$, is given as the solution of the estimating equations (Thas et al., 2012)

$$\sum_{(i,j) \in \mathcal{I}_n} \mathbf{U}_{ij}(\boldsymbol{\beta}) = \mathbf{0}, \quad \mathbf{U}_{ij}(\boldsymbol{\beta}) := \mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) [I(Y_i \preceq Y_j) - g^{-1}(\mathbf{Z}_{ij}^T \boldsymbol{\beta})], \quad (2)$$

with index function $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) = (\partial g^{-1}(\mathbf{Z}_{ij}^T \boldsymbol{\beta}) / \partial \boldsymbol{\beta}) \text{Var}(I(Y_i \preceq Y_j) \mid \mathbf{Z}_{ij})^{-1}$. This choice of index function gives estimating equations similar to those of a generalized linear model. In this paper, however, we often use the identity link function (i.e. $g(u) = u$) and the ordinary least squares (OLS) estimator, resulting from $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) = \mathbf{Z}_{ij}$ in (2).

Despite the mutual independence of the original sample observations, the pseudo-observations are no longer mutually independent. In particular, pseudo-observations $I(Y_i \preceq Y_j)$ and $I(Y_k \preceq Y_l)$ are independent if and only if all indices i, j, k , and l are distinct, i.e. if the pseudo-observations do not share a common response observation. Due to this type of dependence, the sparse correlation distributional theory of Lumley and Mayer-Hamblett (2003) is required for inference.

Thas et al. (2012) showed that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$ as $n \rightarrow \infty$ and they gave a

consistent sandwich estimator of Σ , which is here denoted by $n\hat{\Sigma}_{\hat{\beta}}$ where

$$\hat{\Sigma}_{\hat{\beta}} = \left(\sum_{(i,j) \in \mathcal{I}_n} \frac{\partial \mathbf{U}_{ij}(\hat{\beta})}{\partial \beta^T} \right)^{-1} \left(\sum_{(i,j) \in \mathcal{I}_n} \sum_{(k,l) \in \mathcal{I}_n} \phi_{ijkl} \mathbf{U}_{ij}(\hat{\beta}) \mathbf{U}_{kl}^T(\hat{\beta}) \right) \left(\sum_{(i,j) \in \mathcal{I}_n} \frac{\partial \mathbf{U}_{ij}(\hat{\beta})}{\partial \beta^T} \right)^{-T} \quad (3)$$

with $\phi_{ijkl} = 1$ if $\mathbf{I}(Y_i \preceq Y_j)$ and $\mathbf{I}(Y_k \preceq Y_l)$ are correlated and $\phi_{ijkl} = 0$ otherwise.

For the remainder of the paper we restrict the PIM to factorial designs with one or two factors of interest and possibly one blocking factor or a continuous covariate. The specific definition of the pseudo-observations $\mathbf{I}(Y_i \preceq Y_j) = \mathbf{I}(Y_i < Y_j) + 0.5\mathbf{I}(Y_i = Y_j)$ can deal with ties, but to avoid lengthy formulae we assume no ties are present.

For the factorial design with a single factor and a blocking factor we write $\mathbf{X} = (X, B)$, where X is a factor referring to groups or treatments of interest, and B is a blocking factor. Without loss of generality we say that X takes values $1, \dots, K$, and B takes values $1, \dots, L$. The number of replicates for $X = k$ and $B = l$ is denoted by n_{kl} and the total sample size is given by $N = \sum_{k=1}^K \sum_{l=1}^L n_{kl}$. Let F_{kl} denote the distribution function of Y given $X = k$ and $B = l$. In the absence of blocks, set $B = 1$ and we use the simplified notation n_k for the number of replicates for $X = k$ and F_k for the distribution function of Y given $X = k$. The notation for the two factor setting and the randomized design with covariate is similar, but postponed to Sections 3.3 and 5 respectively.

Sometimes it will be convenient to work with the classical ANOVA notation. Throughout the paper it will be clear from the context which notation is used. In particular, for $\mathbf{X} = (X, B)$, Y_{kl} denotes a random response variable in treatment group $k = 1, \dots, K$ and block $l = 1, \dots, L$. The index l becomes obsolete in the absence of blocks. We use $Y_{\cdot l}$ to denote the random response variable whose distribution is marginalized over the treatment groups, but still conditional on block l . To distinguish between the notation and model as in (1) and the ANOVA form, we refer to the former as the *regression model*, whereas models with the ANOVA notation will be referred to

as the *ANOVA model*. Just like with classical linear regression models, the estimation of the parameters requires that ANOVA models are translated into regression models with dummy regressors for the coding of the factors.

3 THE MARGINAL PIM

3.1 K -sample layout

As a first model we define the *marginal PIM* for the K -sample layout in the absence of blocks. It is marginal in the sense that we only condition on one treatment within the PI, i.e. $P(Y_i \preceq Y_j \mid X_j)$. This PI refers to the distribution of the response of observation j conditional on its regressor, i.e. $Y_j \mid X_j$, and the marginal response distribution of an observation i , i.e. Y_i . In terms of the ANOVA notation and if $X_j = k$, this becomes $P(Y_i \preceq Y_k)$, with Y_k a random response with distribution F_k and Y_i a random response with distribution $F_i = \sum_{k=1}^K \lambda_k F_k$ with $\lambda_k = \lim_{N \rightarrow \infty} n_k/N$ where we assume $\lambda_k > 0$. Consider the marginal PIM in ANOVA form,

$$P(Y_i \preceq Y_k) = \alpha_k. \quad (4)$$

The interpretation of α_k is immediate: it is the probability that a random observation of group k exceeds a random observation of the marginal distribution. The corresponding PIM regression model is obtained upon defining

$$\mathbf{Z}_{ij}^T = \left(I(X_j = 1), \dots, I(X_j = K) \right), \quad (5)$$

for all pairs of regressors (X_i, X_j) . Let $\boldsymbol{\alpha}^T = (\alpha_1, \dots, \alpha_K)$. Model (4) now becomes

$$P(Y_i \preceq Y_j \mid X_j) = \mathbf{Z}_{ij}^T \boldsymbol{\alpha}, \quad (6)$$

with $\mathcal{X}_n = \{(X_i, X_j) \mid i, j = 1, \dots, N\}$, i.e. we consider all N^2 pairs of observations. This model is closely related to the comparison mid-probability index as discussed in

Parzen and Mukhopadhyay (2012). Our model also follows from the nonparametric model of Akritas and Arnold (1994); see Section C of the supplementary material for more details.

Let $\mathbf{1}$ denote the unit vector of length K , \mathbf{A}^- a generalized inverse of a matrix \mathbf{A} and $\hat{\boldsymbol{\alpha}}$ the estimator of $\boldsymbol{\alpha}$ in (6), defined as the solution of the estimating equations (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) = \mathbf{Z}_{ij}$. In Section A of the supplementary material, we show that, under $H_0 : F_1 = \dots = F_K$, $\boldsymbol{\Sigma}_0 := \text{Var}(\hat{\boldsymbol{\alpha}} \mid H_0) = (N+1)(12N)^{-1} \text{diag}(n_1^{-1}, \dots, n_K^{-1}) \mathbf{M}$, with $\mathbf{M} = \mathbf{I} - N^{-1} \text{diag}(n_1, \dots, n_K) \mathbf{1}\mathbf{1}^T$, where \mathbf{I} denotes the $K \times K$ identity matrix. The following theorem establishes the relationship between the marginal PIM and the KW statistic,

$$\text{KW}_s := \frac{12}{N(N+1)} \sum_{k=1}^K n_k \left(\bar{R}_k - \frac{N+1}{2} \right)^2, \quad (7)$$

where \bar{R}_k denotes the average rank of the sample observations in group $X = k$, for which the ranking is performed in the pooled sample. The proof is provided in Section A of the supplementary material.

Theorem 1 (Kruskal–Wallis). *Let $\hat{\boldsymbol{\alpha}}$ denote the estimator of $\boldsymbol{\alpha}$ in (6), defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) = \mathbf{Z}_{ij}$, and let $\boldsymbol{\Sigma}_0$ denote its covariance matrix under the null hypothesis of equal distributions, then*

$$\left(\hat{\boldsymbol{\alpha}} - \frac{1}{2} \mathbf{1} \right)^T \boldsymbol{\Sigma}_0^{-1} \left(\hat{\boldsymbol{\alpha}} - \frac{1}{2} \mathbf{1} \right) = \text{KW}_s. \quad (8)$$

Observe that we denote the KW test statistic as KW_s . The subscript s is used to indicate that this is a *score-type* test, in the sense that the covariance matrix $\boldsymbol{\Sigma}_0$ in (8) is only valid under H_0 . The PIM theory provides a sandwich estimator of the covariance matrix, given by (3), which is also consistent under the alternative. For notational convenience we drop the subscript $\hat{\boldsymbol{\beta}}$ in (3) and write $\hat{\boldsymbol{\Sigma}}$. It is thus straightforward to construct a *Wald-type* KW test by replacing $\boldsymbol{\Sigma}_0$ by $\hat{\boldsymbol{\Sigma}}$ in (8).

These Wald-type tests can e.g. be used to construct exact and asymptotically robust permutation tests for testing $H_0 : P(Y_i \preceq Y_k) = 0.5 \forall k$, which is less restrictive than $H_0 : F_1 = \dots = F_K$; see [Chung and Romano \(2013\)](#). Since the marginal PIM parameters are interpretable effect sizes, it can be of interest to use $\hat{\Sigma}$ for constructing confidence intervals for these parameters.

Note that instead of considering Wald-type statistics for testing $H_0 : P(Y_i \preceq Y_k) = 0.5 \forall k$, better small sample performance may be obtained by constructing ANOVA-type statistics; see [Brunner and Puri \(2001\)](#). Alternatively, the F-type statistics of [Akritas et al. \(2009\)](#) can also be considered.

3.2 Randomized complete block design

The marginal PIM can be extended to block designs. In ANOVA notation this becomes

$$P(Y_{.l} \preceq Y_{.kl}) = \alpha_k, \quad (9)$$

where $k = 1, \dots, K$ refers to the treatment group and $l = 1, \dots, L$ to the block. The interpretation of α_k is immediate: it is the probability that a random observation of group k exceeds a random observation of the marginal distribution *within the same block*. Let \mathbf{Z}_{ij} as in (5) and $\boldsymbol{\alpha}$ as before. Model (9) in regression notation becomes

$$P(Y_i \preceq Y_j \mid B_i, X_j, B_j) = \mathbf{Z}_{ij}^T \boldsymbol{\alpha}, \quad (10)$$

which is now only defined for $(\mathbf{X}_i, \mathbf{X}_j) \in \mathcal{X}_n = \{(\mathbf{X}_i, \mathbf{X}_j) \mid B_i = B_j, i, j = 1, \dots, N\}$, i.e. we restrict the PI to comparisons within blocks. The blocking does not result in extra parameters in the model, but it affects the estimating equations through a restriction on the pseudo-observations to be included.

Consider a randomized complete block (RCB) design for which each treatment-block combination has a constant non-zero number of replicates, i.e. $n_{ij} = n =$

$N/(KL) \geq 1$. For testing the null hypothesis $H_0 : F_{1l} = \dots = F_{Kl}$ ($l = 1, \dots, L$), the MS test (Mack and Skillings, 1980) is an appropriate test. The test statistic is given by

$$MS_s := \frac{12}{K(N+L)} \sum_{k=1}^K \left(\bar{R}_k - \frac{N+L}{2} \right)^2, \quad (11)$$

where $\bar{R}_k = n^{-1} \sum_{l=1}^L \sum_{j=1}^n R_{klj}$ with R_{klj} denoting the rank of the j^{th} replicate of the response observation of treatment k in block l , for which the ranking is performed within blocks.

In Section A of the supplementary material we show that the covariance matrix of the vector $\hat{\alpha}$ under H_0 may be written as $\Sigma_0 := \text{Var}(\hat{\alpha} | H_0) = (nK + 1)(12KLn^2)^{-1}\mathbf{M}$ where $\mathbf{M} = \mathbf{I} - K^{-1}\mathbf{1}\mathbf{1}^T$. The following theorem establishes the link between the marginal PIM and the MS test. The proof and some further details can be found in Section A of the supplementary material.

Theorem 2 (Mack–Skillings). *For a RCB design, with $\hat{\alpha}$ the estimator of α in (10) defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \beta) = \mathbf{Z}_{ij}$, and with Σ_0 its covariance matrix under the null hypothesis of equal distributions within blocks, it holds that*

$$\left(\hat{\alpha} - \frac{1}{2}\mathbf{1} \right)^T \Sigma_0^{-1} \left(\hat{\alpha} - \frac{1}{2}\mathbf{1} \right) = MS_s.$$

The Friedman test is also embedded in the marginal PIM, for it is a special case of the MS test with $n = 1$, i.e. exactly one observation in each treatment-block combination.

Corollary 1 (Friedman). *If each treatment-block combination in a RCB design has exactly one replicate and if $\hat{\alpha}$ denotes the estimator of α in (10), defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \beta) = \mathbf{Z}_{ij}$ and Σ_0 its covariance matrix under the null hypothesis of equal distributions within blocks, then*

$$\left(\hat{\alpha} - \frac{1}{2}\mathbf{1} \right)^T \Sigma_0^{-1} \left(\hat{\alpha} - \frac{1}{2}\mathbf{1} \right) = \frac{12L}{K(K+1)} \sum_{k=1}^K \left(\bar{R}_k - \frac{K+1}{2} \right)^2, \quad (12)$$

in which the right hand side of the equation is exactly the Friedman rank test statistic.

We refer to the MS and Friedman statistics as MS_s and F_s , respectively, where the subscript s indicates that these are score-type test statistics. Similar as for the KW test, we can construct Wald-type MS and Friedman statistics by replacing Σ_0 in (12) by the sandwich estimator $\hat{\Sigma}$. Furthermore, confidence intervals for the individual effect sizes α_k can also be computed.

3.3 The two-way layout

Based on the established relationship between a PIM and the KW test, we now formulate a marginal PIM to construct a rank test for the two-way layout.

Let $\mathbf{X} = (X_1, X_2)$ where X_1 corresponds to the first factor with K_1 levels and X_2 to the second factor with K_2 levels. When using the ANOVA notation, Y_{kl} denotes a response variable associated with groups $X_1 = k$ and $X_2 = l$ and F_{kl} is the corresponding distribution function. We use the notation $Y_{k.}$ to denote responses of the distribution marginalized over X_2 , i.e. $F_{k.} = \sum_{l=1}^{K_2} \lambda_{kl} F_{kl}$ where $\lambda_{kl} = \lim_{n_{k.} \rightarrow \infty} n_{kl}/n_{k.}$, with $n_{k.} = \sum_{l=1}^{K_2} n_{kl}$. Similar for $Y_{.l}$ with distribution $F_{.l} = \sum_{k=1}^{K_1} \lambda_{kl}^* F_{kl}$ where $\lambda_{kl}^* = \lim_{n_{.l} \rightarrow \infty} n_{kl}/n_{.l}$ with $n_{.l} = \sum_{k=1}^{K_1} n_{kl}$ and $Y_{..}$ with distribution $F_{..} = \sum_{k=1}^{K_1} \sum_{l=1}^{K_2} \lambda_{kl}^{**} F_{kl}$ where $\lambda_{kl}^{**} = \lim_{N \rightarrow \infty} n_{kl}/N$. All λ 's are assumed to be strict positive.

A natural extension of (4) to the two-way layout is given by

$$P(Y_{..} \preceq Y_{kl}) = \mu + \alpha_k + \beta_l + (\alpha\beta)_{kl}, \quad (13)$$

with restrictions $\sum_k \alpha_k = 0$, $\sum_l \beta_l = 0$, $\sum_l (\alpha\beta)_{kl} = 0$, and $\sum_k (\alpha\beta)_{kl} = 0$. The parametrization of (13) can be justified upon using the decomposition of F_{kl} as in Akritas and Arnold (1994), i.e.

$$F_{kl}(y) = M(y) + A_k(y) + B_l(y) + C_{kl}(y),$$

where $\sum_k A_k(y) = 0$, $\sum_l B_l(y) = 0$, $\sum_k C_{kl}(y) = 0$, and $\sum_l C_{kl}(y) = 0$ for all y . The relationship with (13) is established by noting that $M(y) = F_{..}(y)$, $P(Y_{..} \preceq Y_{kl}) = \int M(y) dF_{kl}(y)$, and defining $\mu := \int M(y) dM(y) \equiv 0.5$, $\alpha_k := \int M(y) dA_k(y)$, $\beta_l := \int M(y) dB_l(y)$, and $(\alpha\beta)_{kl} := \int M(y) dC_{kl}(y)$; see section C of the supplementary material for more information.

Test statistics for testing main and interaction effects can now be constructed in a similar fashion as in Section 3.1. For example, if $(\alpha\beta)$ denotes the vector with elements $(\alpha\beta)_{ij}$ and $(\widehat{\alpha\beta})_{ij}$ the estimator of $(\alpha\beta)_{ij}$ in (13), defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \beta) = \mathbf{Z}_{ij}$, a test for interaction can be based on the statistic $(\widehat{\alpha\beta})^T \Sigma_0^{-1} (\widehat{\alpha\beta})$, which asymptotically follows a chi-squared distribution with $(K_1 - 1)(K_2 - 1)$ degrees of freedom under $H_0 : F_{kl} = F_{..}$ for all $k = 1, \dots, K_1$ and $l = 1, \dots, K_2$. The matrix Σ_0 is the variance-covariance matrix of $(\widehat{\alpha\beta})$, evaluated under H_0 . Since this null hypothesis is often too restrictive to test for interaction, as it also implies the absence of main effects, a Wald-type test can be constructed, using the statistic $(\widehat{\alpha\beta})^T \hat{\Sigma}^{-1} (\widehat{\alpha\beta})$ with $\hat{\Sigma}$ given by (3). This statistic also follows asymptotically a chi-squared distribution with $(K_1 - 1)(K_2 - 1)$ degrees of freedom, but now under the less restrictive $H_0 : (\alpha\beta)_{kl} = 0$ for all $k = 1, \dots, K_1$ and $l = 1, \dots, K_2$. This test statistic is closely related to work of Brunner and Puri (2001). The interpretation of $(\alpha\beta)_{kl}$ can be derived from $(\alpha\beta)_{kl} = [P(Y_{..} \preceq Y_{kl}) - P(Y_{..} \preceq Y_{.l})] - [P(Y_{..} \preceq Y_{k.}) - P(Y_{..} \preceq Y_{..})]$.

In a similar fashion as in Section 3.2, PIM (13) can be extended to block-designs. In Section C of the supplementary material we also show how PIM (4) can be extended with a continuous regressor.

4 THE PAIRWISE PIM

In the absence of blocks the marginal PIM for the K -sample layout is associated with rank tests which are based on the joint ranking. The model refers to the comparison of the marginal response with a response in a particular treatment group, i.e. $P(Y_i \preceq Y_k)$. In this section we propose a PIM that models pairwise comparisons between treatment groups. In particular, for the K -sample design we propose the PIM (ANOVA notation)

$$P(Y_k \preceq Y_l) = \alpha_{kl}. \quad (14)$$

The parameter α_{kl} thus gives the probability that a random observation of group l exceeds a random observation of group k . The regression PIM follows from defining

$$\mathbf{Z}_{ij}^T = [I(X_i = 1)I(X_j = 2), I(X_i = 1)I(X_j = 3), \dots, I(X_i = K-1)I(X_j = K)], \quad (15)$$

and $\boldsymbol{\alpha}$ the vector with the corresponding elements α_{kl} . The pairwise PIM then becomes (regression notation)

$$P(Y_i \preceq Y_j \mid X_i, X_j) = \mathbf{Z}_{ij}^T \boldsymbol{\alpha}, \quad (16)$$

with $\mathcal{X}_n = \{(\mathbf{X}_i, \mathbf{X}_j) \mid X_i < X_j, i, j = 1 \dots, N\}$, i.e. we restrict the PI to all unique treatment combinations.

4.1 Two-sample layout

In the following theorem we establish the relationship between the pairwise PIM and the WMW test; see Section B of the supplementary material for the proof.

Theorem 3 (Wilcoxon–Mann–Whitney). *For the two-sample layout, let $\hat{\alpha}_{12}$ denote the estimator associated with PIM (16) defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) =$*

\mathbf{Z}_{ij} , and let σ_0^2 denote its variance under the null hypothesis of equal distributions, then

$$\frac{\hat{\alpha}_{12} - 0.5}{\sigma_0} = \frac{\sum_{\{i|X_i=1\}} \sum_{\{j|X_j=2\}} \mathbf{I}(Y_i \preceq Y_j) - n_1 n_2 / 2}{\sqrt{[n_1 n_2 (n_1 + n_2 + 1)] / 12}}, \quad (17)$$

in which the right hand side of the equation is exactly the standardized WMW statistic.

4.2 Three-sample layout

In this section we establish the relationship between the pairwise PIM and the rank test of [Brown and Hettmansperger \(2002\)](#) for the three-sample layout. [Brown and Hettmansperger \(2002\)](#) showed that the KW test cannot test for intransitivity and therefore they proposed to extend the KW test statistic as follows

$$\text{BH}_s := \text{KW}_s + \frac{3n_1 n_2 n_3}{N} \left(\frac{T_{12}}{n_1 n_2} + \frac{T_{23}}{n_2 n_3} + \frac{T_{31}}{n_3 n_1} \right)^2, \quad (18)$$

where $T_{kl} = \sum_{\{i|X_i=k\}} \sum_{\{j|X_j=l\}} [2\mathbf{I}(Y_i \preceq Y_j) - 1]$. A large value of the second component of BH_s suggests intransitivity. Let $\hat{\boldsymbol{\alpha}}^T = (\hat{\alpha}_{12}, \hat{\alpha}_{23}, \hat{\alpha}_{13})$ denote the estimators of the parameters in the PIM (16), and let $\boldsymbol{\Sigma}_0$ denote the covariance matrix of $\hat{\boldsymbol{\alpha}}$ under the null hypothesis $H_0 : F_1 = F_2 = F_3$. The following theorem establishes the relationship between the pairwise PIM and the BH statistic. The proof is provided in Section B of the supplementary material.

Theorem 4 (Brown–Hettmansperger). *Let $\hat{\boldsymbol{\alpha}}$ denote the estimator associated with PIM (16) defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) = \mathbf{Z}_{ij}$ and let $\boldsymbol{\Sigma}_0$ denote its covariance matrix under the null hypothesis of equal distributions, then*

$$\left(\hat{\boldsymbol{\alpha}} - \frac{1}{2} \mathbf{1} \right)^T \boldsymbol{\Sigma}_0^{-1} \left(\hat{\boldsymbol{\alpha}} - \frac{1}{2} \mathbf{1} \right) = \text{BH}_s, \quad (19)$$

with BH_s given by (18).

4.3 Ordered alternatives

Thus far all tests focused on rejecting the null hypothesis in favour of an alternative that states that some particular PI's are not equal to 0.5. However, sometimes more informative alternatives can be of interest. When the K treatments can be ordered (e.g. the dosage of a drug), one can sometimes formulate an alternative for which the response tends to increase or decrease with increasing treatment level. Mann (1945) defined an upward trend as

$$H_1 : \frac{2}{K(K-1)} \sum_{k=1}^{K-1} \sum_{l=k+1}^K P(Y_i \preceq Y_j \mid X_i = l, X_j = k) > \frac{1}{2}.$$

Under the location shift model $F_1(y - \tau_1) = \dots = F_K(y - \tau_K)$, H_1 simplifies to the ordered alternative $\tau_1 \leq \dots \leq \tau_K$, with at least one strict inequality. The Jonckheere–Terpstra (JT) test (Jonckheere, 1954; Terpstra, 1952) is consistent against H_1 and its test statistic is given by

$$JT_s := \sigma_{JT}^{-1} \left(\sum_{k=1}^{K-1} \sum_{l=k+1}^K \sum_{\{i \mid X_i=k\}} \sum_{\{j \mid X_j=l\}} I(Y_i \preceq Y_j) - \mu_{JT} \right), \quad (20)$$

where $\mu_{JT} = (N^2 - \sum_{j=1}^K n_j^2)/4$ and $\sigma_{JT}^2 = [N^2(2N+3) - \sum_{j=1}^K n_j^2(2n_j+3)]/72$.

The JT test can also be obtained from the PIM (regression notation)

$$P(Y_i \preceq Y_j \mid X_i, X_j) = \frac{1}{2} + \alpha Z_{ij}, \quad (21)$$

where $Z_{ij} = I(X_i < X_j) - I(X_i > X_j)$ and $\mathcal{X}_n = \{(X_i, X_j) \mid i, j = 1, \dots, N\}$. The interpretation of α comes from $\alpha = P(Y_i \preceq Y_j \mid X_i < X_j) - 0.5$, i.e. the probability that an outcome of a higher factor level exceeds an outcome of a lower factor level, reduced with 0.5. Equivalently, $\alpha = 0.5 - P(Y_i \preceq Y_j \mid X_j < X_i)$. Note that the offset 0.5 is a consequence of the definition of Z_{ij} , which must guarantee that $P(Y_i \preceq Y_j \mid X_i = X_j) = 0.5$. Let $\hat{\alpha}$ denote the OLS estimator associated with PIM (21), then its variance under the null hypothesis of equal distributions, say σ_0^2 , can be obtained in a similar fashion as in Section 3.

The hypothesis of interest $H_0 : F_1 = \dots = F_K$ is tested versus the alternative $H_1 : \alpha \neq 0$ which is equivalent to $H_1 : P(Y_i \preceq Y_j \mid X_i < X_j) \neq 0.5$. We prefer to use the regression notation here, because the factor acts as an integer-valued regressor. The following theorem establishes the relationship between PIM (21) and the JT test. The proof is provided in Section B of the supplementary material.

Theorem 5 (Jonckheere–Terpstra). *Let $\hat{\alpha}$ denote the OLS estimator associated with PIM (21) defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) = \mathbf{Z}_{ij}$ and let σ_0^2 denote its variance under the null hypothesis of equal distributions, then $\hat{\alpha}/\sigma_0 = JT_s$, with JT_s given by (20).*

In the Section B of the supplementary material it also demonstrated how a similar connection can be made to the Mack–Wolfe test (Mack and Wolfe, 1981) for umbrella alternatives.

Similar as for the marginal PIM, the pairwise PIM can be extended to block designs by limiting the comparisons in \mathcal{X}_n to pseudo-observations within a block.

5 RANDOMIZED DESIGNS WITH A COVARIATE

Consider a randomized design where a binary factor X is randomized over experimental units and it is of interest to test the null hypothesis $H_0 : F_1 = F_2$, with F_i the conditional distribution function of Y given $X = i$. In addition to X we also measure a continuous covariate Z which is associated with the response variable Y but independent of X due to the randomization. As an illustration, consider a setting where one wants to assess the difference between two fertilizers (X) in terms of the yield Y and that fields form the experimental units. If one randomly allocates a fertilizer to a field, Z can be e.g. the hours of sunshine or the amount of rainfall for each field. The random allocation of X assures the independence between X and Z , while the yield

Y can still depend on Z . We will construct a rank test that takes the information in Z into account and which can be more powerful than the WMW test.

More specifically, we propose the PIM with identity link

$$P(Y_i \preceq Y_j \mid \mathbf{X}_i, \mathbf{X}_j) = \beta + \gamma[f(Z_j) - f(Z_i)], \quad \mathcal{X}_n = \{\mathbf{X}_i, \mathbf{X}_j \mid X_i = 1, X_j = 2\}, \quad (22)$$

with $\mathbf{X} = (X, Z)$ and $f(\cdot)$ a known function.

The following theorem provides the distributional properties of the estimator of β in (22) as well as the interpretation.

Theorem 6. *Let $\hat{\beta}$ denote the estimator of the PIM (22) defined as the solution of (2) with $\mathbf{W}(\mathbf{Z}_{ij}, \boldsymbol{\beta}) = \mathbf{Z}_{ij}$. Assume that X and Z are independent, and $n_1/N \rightarrow \lambda \in (0, 1)$ as $N \rightarrow \infty$. Then, as $N \rightarrow \infty$,*

$$\sqrt{N} \frac{\hat{\beta} - P(Y_i \preceq Y_j \mid X_i = 1, X_j = 2)}{\sigma} \xrightarrow{d} N(0, 1), \quad (23)$$

where

$$\begin{aligned} \sigma^2 &= \frac{1}{\lambda} \text{Var}(F_2(Y_i) - \gamma f(Z_i) \mid X_i = 1) + \frac{1}{1 - \lambda} \text{Var}(F_1(Y_i) - \gamma f(Z_i) \mid X_i = 2), \\ \gamma &= \frac{\text{Cov}(I(Y_i \preceq Y_j), f(Z_j) - f(Z_i) \mid X_i = 1, X_j = 2)}{\text{Var}(f(Z_j) - f(Z_i) \mid X_i = 1, X_j = 2)}, \end{aligned}$$

where F_i denotes the conditional distribution of Y given $X = i$ ($i = 1, 2$).

The proof can be found in Section D of the supplementary material. A Wald-type test statistic can now be constructed as $(\hat{\beta} - 0.5)/\hat{\sigma}$, where $\hat{\sigma}$ is the square root of the first diagonal element in (3) and which is a consistent estimator of σ . From Theorem 6 it follows that $\hat{\beta}$ is a consistent estimator for the probability $P(Y_i \preceq Y_j \mid X_i = 1, X_j = 2)$, i.e. the probability that a randomly selected response for which $X = 2$ exceeds a randomly selected response for which $X = 1$. Note that this probability does not depend on Z and corresponds to the population parameter tested for by the WMW test. Indeed, for the WMW statistic

$\hat{\alpha} := (n_1 n_2)^{-1} \sum_{\{i:X=1\}} \sum_{\{j:X=2\}} \mathbf{I}(Y_i \preceq Y_j)$ it holds that

$$\sqrt{N} \frac{[\hat{\alpha} - \mathbf{P}(Y_i \preceq Y_j \mid X_i = 1, X_j = 2)]}{\sigma^*} \xrightarrow{d} \mathbf{N}(0, 1), \quad (24)$$

where

$$\sigma^{*2} = \frac{1}{\lambda} \text{Var}(F_2(Y_i) \mid X_i = 1) + \frac{1}{1-\lambda} \text{Var}(F_1(Y_i) \mid X_i = 2),$$

see e.g. [van der Vaart \(1998\)](#), p. 197. Since $\hat{\beta}$ and $\hat{\alpha}$ both converge in probability to $\mathbf{P}(Y_i \preceq Y_j \mid X_i = 1, X_j = 2)$, it follows that the asymptotic relative efficiency (ARE) of the test based on $\hat{\beta}$ versus the test based on $\hat{\alpha}$ (i.e. the WMW test) only depends on the ratio σ^{*2}/σ^2 evaluated under the null hypothesis $H_0 : F_1 = F_2$.

To compute the ARE's we consider the working model and sequence of alternatives

$$Y_i = \mu + \frac{\delta_1}{\sqrt{N}} X_i + \delta_2 f(Z_i) + \varepsilon_i, \quad \delta_1 > 0, \quad Z \stackrel{d}{=} \mathbf{N}(0, \sigma_Z^2), \quad (25)$$

for several choices of δ_2 , σ_Z^2 , and error distribution, scaled to have expectation zero and variance one. More specifically, we choose the normal, logistic, and lognormal distributions for ε . We consider three functions: $f(Z) = Z$, $f(Z) = Z^2$ and $f(Z) = \sin(Z)$ and set $\mu = 0$ without loss of generality. Table 1 gives these ARE's for testing $H_0 : \delta_1 = 0$ versus $H_1 : \delta_1 > 0$ for a balanced design ($\lambda = 0.5$) for which σ^2 in (23) is approximated using 10^7 Monte-Carlo simulations. An ARE greater than one means that the new test is asymptotically more efficient than the WMW test. In the absence of a covariate effect ($\delta_2 = 0$) both tests have the same efficiency. When $\delta_2 > 0$ the new test becomes more efficient than the WMW test when the magnitude of the covariate effect increases. Furthermore, in the presence of a covariate effect, if the variance of Z increases, the ARE increases as well. These conclusions hold for all three choice of $f(\cdot)$. When the functional form of Z in (25) and (22) differs, $\hat{\beta}$ remains a consistent estimator for $\mathbf{P}(Y_i \preceq Y_j \mid X_i = 1, X_j = 2)$, but the asymptotic efficiency changes. We refer to Section D of the supplementary material for more ARE calculations.

Table 1: Asymptotic relative efficiencies of the new test based on $\hat{\beta}$ versus the WMW test based on $\hat{\alpha}$, for several choices of δ_2 , σ_Z^2 , $f(\cdot)$ and error distribution in (25), where N denotes the normal, L the logistic and LN the lognormal distribution.

		$f(Z) = Z$			$f(Z) = Z^2$			$f(Z) = \sin(Z)$		
Error distribution		N	L	LN	N	L	LN	N	L	LN
δ_2	σ_Z^2									
0	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.25	1.24	1.26	1.75	1.10	1.11	1.27	1.19	1.21	1.67
2	0.25	1.91	1.97	2.94	1.28	1.30	1.53	1.77	1.84	2.83
0	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	1	1.91	1.98	2.95	1.63	1.64	1.86	1.44	1.49	2.25
2	1	4.23	4.31	5.78	2.03	2.04	2.18	2.90	2.98	4.53
0	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	4	4.24	4.31	5.79	2.33	2.33	2.39	1.52	1.57	2.39
2	4	9.87	9.92	11.14	2.47	2.48	2.49	3.24	3.31	4.82

6 EXAMPLE

To illustrate how a PIM regression representation of rank tests can improve the understanding of the data, we consider a subset of the chick weight dataset as described in [Crowder and Hand \(1990\)](#). Chicks are randomly allocated to one of four diets: a normal diet (referred to as diet 1) or one of three specific diets with respectively 10%, 20% or 40% protein replacement (referred to as diets 2, 3 or 4, respectively). The weights (in gram) of the chicks are measured on alternate days for the first three weeks after birth, but we only look at the weight measured at day 6. Figure 1 shows the boxplots of the weights at day 6 for the four diets.

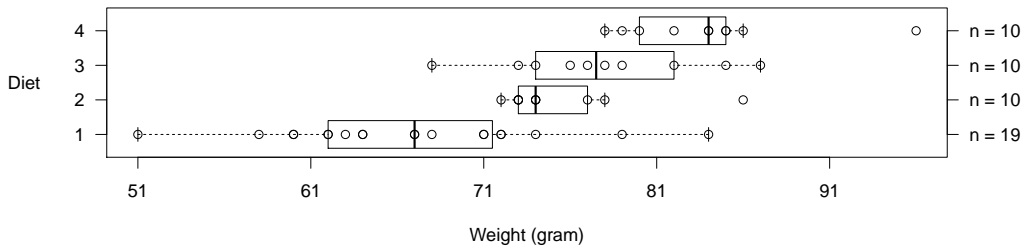


Figure 1: Box and strip-plots of the chick weights for the four diets. The right axis shows the sample sizes.

The Kruskal-Wallis rank test rejects the null-hypothesis of equal weight distributions at the 5% level of significance ($KW_s = 28.3$ and $p < 0.0001$). Since the KW test is related to the marginal PIM (4), we estimate the PIM regression parameters. It follows that $\hat{\alpha}_1 = 0.25$ with a 95% confidence interval $[0.16, 0.33]$, $\hat{\alpha}_2 = 0.52$ $[0.40, 0.65]$, $\hat{\alpha}_3 = 0.63$ $[0.50, 0.76]$ and $\hat{\alpha}_4 = 0.83$ $[0.75, 0.90]$. For example, a randomly selected chick that received diet 4 has a 83% chance to have a higher weight after 6 days as compared to a randomly selected chick from the pooled sample.

For an ordered alternative, ordered according to the percentage of protein replacement, the JT-test rejects the null-hypothesis of equal weight distributions ($JT_s = 5.79$ and $p < 0.0001$). The α parameter of PIM (21) is estimated as 0.37 [0.29, 0.44]. Consequently $\hat{P}(\text{Weight}_i \preceq \text{Weight}_j \mid \text{Diet}_i < \text{Diet}_j) = 0.5 + 0.37 = 0.87$ [0.79, 0.94]: there is an 87% chance that a chick receiving a diet with more protein replacement will have a higher weight than a chick receiving a diet with less protein replacement.

The pairwise PIM (14) can be used to compare the diets two-by-two. For example, when comparing diet 1 with 4 the probabilistic index is estimated by $\hat{\alpha}_{14} = 0.96$ [0.92, 1]: there is a 96% chance that a chick receiving diet 4 will have a higher weight than a chick receiving diet 1. Since the birth weights of the chicks are also measured and since these are independent of the diets, the probabilistic index can also be estimated with PIM (22) for some function $f(\cdot)$, accounting for the weights at baseline. For a linear $f(\cdot)$ it follows that $\hat{\beta}_{14} = 0.98$, providing a more narrow 95% confidence interval [0.95, 1] as compared to the results of the PIM (14). It is illustrated in Section E of the supplementary material how these analyses can be conducted using the R package `pim` (R, 2014).

7 DISCUSSION

We have illustrated how two parametrizations of a PIM can lead to rank tests for factorial designs. Based on the marginal PIM parametrization, the Kruskal–Wallis (KW) test for a K -sample layout and the Mack–Skillings and Friedman tests for a randomized complete block design, arise naturally. The pairwise PIM results in the Wilcoxon–Mann–Whitney (WMW) test for the two-sample layout and the Brown–Hettmansperger test for the three-sample layout. The pairwise PIM also generates the Jonckheere–Terpstra and the Mack–Wolfe tests. All these rank tests are score-type tests in the sense that the variances, which are required for standardizing the test

statistics, are obtained under the null hypothesis. The PIM theory, however, provides a sandwich estimator of the variance which is also consistent under the alternative, thus allowing to construct Wald-type versions of these rank tests, as well as confidence intervals for the effect sizes.

The PIM connection with rank tests may also be valuable for generating rank tests for more complicated designs, in which, for example, covariates, blocking factors or stratification factors are present. The classical rank tests are often referred to as nonparametric tests, but this term may be misleading. Apart from some very simple settings (e.g. K -sample layout) rank tests relate to parameters of a semiparametric model which expresses restrictions on the distribution of the response variable. In this paper we have demonstrated that the PIM is a natural model for rank tests. The potential of the PIM as a rank test framework is illustrated in Section 6 by analyzing an example dataset.

[Akritas and Arnold \(1994\)](#) proposed another model for which they developed rank tests, which, however, do not generally reduce to the classical tests. Their methodology was extended to several designs and to the inclusion of covariates ([Akritas et al., 1997, 2000](#); [Brunner and Puri, 2002](#); [Tsangari and Akritas, 2004](#)). Their test statistics are rank-transform statistics, in the sense that they are functions of the rank-transformed response observations. Although their methods also rely on a semi-parametric model, they cannot always estimate all terms in their model ([Tsangari and Akritas, 2004](#)). At this point it is also interesting to mention that the simple rank-transform methods of [Conover and Iman \(1981\)](#) and [Hora and Conover \(1984\)](#) do not always relate clearly to a statistical model. The method consists in transforming the response observations to their ranks and subsequently using these transformed observations in parametric methods. For example, the two-sample t -test and the one-way ANOVA F -test applied to the rank-transformed data gives the WMW and the

KW test, respectively. However, for more complicated designs Akritas (1990) showed that the parametric statistical model does no longer hold after the transformation. For example, the two-way ANOVA without interaction implies additivity of the effects on the mean response, but this additivity is lost after the transformation. He made the connection without explicitly referring to the probabilistic index. In particular, upon using asymptotic arguments, he replaced the rank-transformed response of Y_i with $nF(Y_i)$, with F the marginal distribution function of the response. When the response is continuous, the original parametric model for $E(Y_i \mid \mathbf{X}_i)$ becomes $E(nF(Y_i) \mid \mathbf{X}_i) = nP(Y \leq Y_i \mid \mathbf{X}_i)$, which resembles our marginal PIM. The additivity of the effects on $E(Y_i \mid \mathbf{X}_i)$ thus becomes additive in the marginal PIM with identity link. To some extent, our method may also be seen as a two-stage approach in which first the sample observations (Y_i, Y_j) are transformed to pseudo-observations $I(Y_i \leq Y_j)$ which are subsequently used as response observations in a linear regression model. By restricting the set of pseudo-observations to comparisons within blocks, block designs can be analyzed. However, despite this apparently simple trick, we do not encourage statisticians to look at it this way. Instead we prefer to interpret our approach within a genuine semiparametric modelling framework. This will help in ensuring correct interpretation and reporting of the data analysis results.

8 SUPPLEMENTARY MATERIAL

supplementary material This file contains proofs, additional results on the relationship with the methods of Akritas and his colleagues, additional results on the randomized design with a covariate and an illustration in R upon using the package `pim`.

References

- Acion, L., Peterson, J., Temple, S., and Arndt, S. (2006), “Probabilistic Index: an Intuitive Non-parametric Approach to Measuring the Size of Treatment Effects,” *Statistics in Medicine*, 25, 591–602.
- Akritis, M.G. (1990), “The Rank Transform Method in Some Two Factor Designs,” *Journal of the American Statistical Association*, 85, 73–78.
- Akritis, M.G., and Arnold, S.F. (1994), “Fully Nonparametric Hypotheses for Factorial Designs I: Multivariate Repeated Measures Designs,” *Journal of the American Statistical Association*, 89, 336–343.
- Akritis, M.G., Arnold, S.F., and Brunner, E. (1997), “Nonparametric Hypotheses and Rank Statistics for Unbalanced Factorial Designs,” *Journal of the American Statistical Association*, 92, 258–265.
- Akritis, M.G., Arnold, S.F., and Du, Y. (2000), “Nonparametric Models and Methods for Nonlinear Analysis of Covariance,” *Biometrika*, 87, 507–526.
- Akritis, M.G., Stavropoulos, A., and Caroni, C. (2009), “Asymptotic Theory of Weighted F-statistics Based on Ranks,” *Journal of Nonparametric Statistics*, 21, 177–191.
- Brown, B.M., and Hettmansperger T.P. (2002), “Kruskal–Wallis, Multiple Comparisons and Efron Dice,” *Australian and New Zealand Journal of Statistics*, 44, 427–438.
- Brunner, E., and Puri, M.L. (2001), “Nonparametric Methods in Factorial Designs,” *Statistical Papers*, 42, 1–52.

- Brunner, E., and Puri, M.L. (2002), “A Class of Rank-Score Tests in Factorial Designs,” *Journal of Statistical Planning and Inference*, 103, 331–360.
- Chung, E.Y., and Romano, J.P. (2013), “Exact and Asymptotically Robust Permutation Tests,” *The Annals of Statistics*, 41, 484–507.
- Conover, W.J., and Iman, R.L. (1981), “Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics,” *American Statistician*, 35, 124–133.
- Crowder, M., and Hand, D. (1990), *Analysis of Repeated Measures*, Chapman and Hal.
- Friedman, M. (1937), “The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance,” *Journal of the American Statistical Association*, 32, 675–701.
- Hajek, K., Sidak, Z., and Sen, P. (1999), *Theory of Rank Tests*, Academic Press – San Diego, USA.
- Hollander, M., and Wolfe, D. (1999), *Nonparametric Statistical Methods*, Wiley – New York, USA.
- Hora, S.C., and Conover, W.J. (1984), “The F Statistic in the Two-Way Layout with Rank-Score Transformed Data,” *Journal of the American Statistical Association*, 79, 668–673.
- Jonckheere, A. R. (1954), “A Distribution-Free k -Sample Test Against Ordered Alternatives,” *Biometrika*, 41, 133–145.
- Kruskal, W., and Wallis, W. (1952), “Use of Ranks in One-Criterion Variance Analysis,” *Journal of the American Statistical Association*, 47, 583–621.

- Lehmann, E.L. (2005), *Nonparametrics. Statistical Methods Based on Ranks*, Prentice Hall – Upper Saddle River, NJ, USA.
- Lumley, T., and Mayer-Hamblett, N. (2003), “Asymptotics for Marginal Generalized Linear Models with Sparse Correlations,” *UW Biostatistics Working Paper Series*, University of Washington, 207.
- Mack, G.A., and Skillings, J.H. (1980), “A Friedman-type Rank Test for Main Effects in a Two-Factor ANOVA,” *Journal of the American Statistical Association*, 75, 947–951.
- Mack, G.A., and Wolfe, D.A. (1981), “K-sample Rank Tests for Umbrella Alternatives,” *Journal of the American Statistical Association*, 76, 175–181.
- Mann, H.B. (1945), “Non-parametric Tests Against Trend,” *Econometrica*, 13, 254–259.
- Mann, H., and Whitney, D. (1947), “On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other,” *Annals of Mathematical Statistics*, 18, 50–60.
- R Core Team (2014), “R: A Language and Environment for Statistical Computing,” R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Parzen, E., and Mukhopadhyay, S. (2012), “Discussion of “Probabilistic Index Models” by Thas et al.,” *Journal of the Royal Statistical Society: Series B*, 74, 623–671.
- Skillings, J.H., and Mack, G.A. (1981), “On the Use of a Friedman-type Statistic in Balanced and Unbalanced Block Designs,” *Technometrics*, 23, 171–177.

- Terpstra, T.J. (1952), “The Asymptotic Normality and Consistency of Kendall’s Test Against Trend, When Ties are Present in One Ranking,” *Indagationes Mathematicae*, 14, 327–333.
- Thas, O. (2009), *Comparing Distributions*, Springer – New York, USA.
- Thas, O., De Neve, J., Clement, L., and Ottoy, JP. (2012), “Probabilistic Index Models (with Discussion),” *Journal of the Royal Statistical Society: Series B*, 74, 623–671.
- Tsangari, H., and Akritas, M.G. (2004), “Nonparametric ANCOVA with Two and Three Covariates,” *Journal of Multivariate Analysis*, 88, 298–319.
- van der Vaart, A. (1998), *Asymptotic Statistics*, Cambridge University Press.
- Wilcoxon, F. (1945), “Individual Comparisons by Ranking Methods,” *Biometrics Bulletin*, 1, 80–83.